International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 9 Issue 1, Jan-Feb 2025 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Breast Cancer Image Classification using GLCM Features based Convolutional Neural Network

Ganesh Chandra[1], Joy Bhattacharji[2], Prof. Anshul Jain[3]

[1]M Tech Scholar, [2,3]Assistant Professor,
[1,2]Department of CSE, TIT-A, Bhopal, Madhya Pradesh, India
[3]Department of CSE, TIT, Bhopal, Madhya Pradesh, India

ABSTRACT
Breast cancer remains one of the leading causes of mortality among women worldwide. Early and accurate diagnosis through medical imaging can significantly improve patient outcomes. Deep learning, particularly Convolutional Neural Networks (CNNs), has shown promising results in classifying medical images with high accuracy. However, the performance of these models often depends on appropriate data preprocessing techniques. This paper investigates the efficacy of using Min-Max Normalization combined with a CNN-based architecture to classify breast cancer images. Experimental results demonstrate that applying Min-Max Normalization prior to training not only enhances model convergence but also improves classification accuracy and robustness. These findings suggest that the proposed approach can provide a reliable diagnostic tool for clinicians in the early detection of breast cancer. This feature matrix is used as input for the pretrained model and convolutional neural network. Pre-trained models such as VGG16 and VGG19 are investigated using the concept of transfer learning. The framework's structure consists of 14 layers in total. In order to optimize the classification accuracy, the hyperparameters are changed. An ideal accuracy of 93.9% is attained by the convolutional neural network architecture that was created.

KEYWORDS: *Convolutional Neural Network, Breast Cancer, VGG16, VGG19, Min-Max Normalization*

How to cite this paper: Ganesh Chandra | Joy Bhattacharji | Prof. Anshul Jain "Breast Cancer Image Classification using GLCM Features based Convolutional Neural Network" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-9 | Issue-1, February 2025, pp.156-165, URL: www.ijtsrd.com/papers/ijtsrd73822.pdf

IJTSRD73822

1. INTRODUCTION
Breast cancer is one of the most prevalent cancers among women and a major global health concern. According to the World Health Organization (WHO), early detection and accurate diagnosis can significantly reduce mortality rates. The use of imaging modalities, such as mammography, ultrasound, and histopathology, is crucial in detecting breast cancer in its early stages. However, manual evaluation of images is time-consuming and prone to human error.

With the advances in Artificial Intelligence (AI) and deep learning, automated systems based on Convolutional Neural Networks (CNNs) have shown remarkable ability to learn intricate patterns from medical images. CNNs, in particular, have demonstrated high performance in tasks such as image recognition, classification, and segmentation. Despite these successes, achieving consistent performance requires careful consideration of data preprocessing.

Normalization of input data is a key step to optimize the training of deep networks. One widely used technique is Min-Max Normalization, which scales each feature to a specified range, often [0, 1] or [−1, 1]. By normalizing inputs, CNNs can converge more quickly and reliably during training, while reducing the risk of getting stuck in poor local minima. Objectives of this study include:

- Investigating the impact of Min-Max Normalization on the classification performance of CNN for breast cancer images.
- Evaluating the model using a public dataset.
- Assessing the classification metrics—accuracy,

precision, recall, and F1-score—to determine the effectiveness of our proposed approach.

2. Literature Review

A recent literature review on deep learning techniques for breast cancer image classification highlights several innovative approaches that have been developed to enhance detection and diagnostic accuracy:

3D CNN Architectures: Studies have utilized 3D convolutional neural networks (CNNs) for the analysis of dynamic contrast-enhanced MRI images, achieving high sensitivity and specificity in detecting malignancies (Adam et al., 2023)

Hybrid Deep Learning Models: Incorporating domain knowledge, some studies have integrated feature learning and adaptive weighting mechanisms into their models, enhancing the interpretability and performance of the diagnosis process in breast cancer detection using MRI and DWI data (Adam et al., 2023).

Comparison with Human Readers: Deep learning models, such as InceptionResNetV2, have been compared to human experts in terms of performance, showing comparable or superior ability to distinguish between benign and malignant lesions (Adam et al., 2023).

Overcoming Data Challenges: Addressing issues like overfitting and imbalanced data, research has focused on developing sophisticated deep learning variants that employ preprocessing, ensemble, and normalization techniques to enhance image quality and classification accuracy (Yusoff et al., 2023).

Convolutional Neural Networks (CNNs): Reviews have evaluated the efficacy of various CNN models in detecting breast cancer, highlighting the strengths and limitations of different architectures and the benefits of integrating transfer learning to reduce the need for large, labeled datasets (Pérez-Núñez et al., 2024)

Feature Fusion and Optimization: Some approaches have focused on optimal feature fusion, using advanced optimization algorithms to select the best features for classification, significantly improving the accuracy of breast cancer classification from ultrasound images (Jabeen et al., 2022).

Automated Systems for Early Detection: There is a strong emphasis on the development of automated systems that can detect breast cancer from ultrasound images efficiently, employing deep learning and feature fusion to achieve high accuracy rates (Jabeen et al., 2022).

These studies collectively indicate the ongoing advancements in the application of deep learning for breast cancer classification, underscoring the potential of these technologies to revolutionize early detection and diagnosis, making them critical components in the fight against breast cancer.

Convolutional Neural Networks in Breast Cancer Classification

Recent literature on convolutional neural networks (CNNs) for breast cancer classification highlights several innovative approaches aimed at improving diagnostic accuracy and model reliability.

Concatenated CNN Models: Some studies have developed concatenated CNN models that integrate multiple pre-trained and from-scratch neural networks. These models aim to increase the accuracy by extracting more comprehensive feature sets from mammography scans. One particular study used three CNN models together with Bayesian optimization for hyperparameter tuning, achieving accuracy rates up to 99.13% in multi-classification scenarios (Alshayeji & Al-Buloushi, 2023).

Meta-Learning and Ensemble Techniques: Other research has focused on combining meta-learning with ensemble techniques using multiple CNNs. This approach helps the models adapt better to new and diverse datasets. Techniques like transfer learning are employed to utilize pre-trained models such as Inception and DenseNet121, enhancing the feature extraction capabilities of the CNNs. Studies reported high effectiveness in classifying breast cancer images as benign or malignant using these advanced methods (Ali et al., 2023).

Optimization and Preprocessing: Additional research has optimized CNNs using methods like histogram equalization for image enhancement before classification. This preprocessing improves the contrast and brightness of mammograms, which can significantly aid in the detection of subtle features indicative of cancer (Shanmugavadivu et al., 2023).

Deep Learning Innovations: Some recent works have explored deep learning innovations like deep bottleneck residual networks, which are designed to improve classification by learning in-depth features from mammogram images. This method could offer a more refined analysis of the images, potentially leading to better classification of breast cancer severity (Jabeen et al., 2024).

These studies collectively advance the application of CNNs in breast cancer classification, offering promising results for improving diagnostic processes and outcomes in medical imaging.

Data Preprocessing and Normalization

Recent studies have explored the effects of different data normalization techniques on various types of data analysis and machine learning models, particularly focusing on Min-Max normalization, Z-score normalization, and using raw images without normalization.

Min-Max Normalization: This method scales data to a fixed range, typically 0 to 1. It is particularly noted for its simplicity and effectiveness in cases where the scales of input features vary widely. Min-Max normalization can be beneficial in enhancing the performance of models like k-NN algorithms, where it has been shown to slightly outperform Z-score normalization in some cases, achieving up to 98% accuracy (Henderi, 2021).

Z-Score Normalization: Also known as standardization, this technique re-scales data to have zero mean and unit variance. This method is especially useful in scenarios where the data does not have a known minimum and maximum value. It is widely used in applications involving deep learning and neural networks due to its ability to maintain the structure of the original data distribution while standardizing the scale (Pranolo et al., 2024).

Using Raw Data without Normalization: Some studies investigate the impact of using raw, unnormalized data. While normalization techniques generally improve the performance of machine learning models by scaling the features to a common range, in some contexts, particularly where the natural variation of the data is meaningful, using raw data might be advantageous. However, this approach is less commonly recommended as it may increase the complexity of the model training phase and affect model convergence and performance. Each of these techniques has its place depending on the nature of the data and the specific requirements of the analytical model being used. For instance, when dealing with highly variable data in terms of scale and distribution, Z-score normalization might be more appropriate. Conversely, for data where maintaining relative distances between values is important, such as in image processing or certain types of time-series data, Min-Max normalization might be more beneficial (Laska & Yolanda, 2024).

3. Proposed Methodology

Dataset Description

The most prevalent malignancy among women worldwide is breast cancer. It affects more than 2.1 million individuals in 2015 alone and makes up 25% of all cancer cases. It begins when breast cells start to proliferate uncontrollably. Usually, these cells develop into tumors that appear on X-rays or are felt as lumps in the breast region. Features calculated from digital photographs of breast mass make up the Breast Cancer Dataset. For every cell nucleus present, it has measurements for properties including radius, texture, perimeter, area, smoothness, compactness, concavity, concave spots, symmetry, and fractal dimension. The main purpose of this dataset is to distinguish between benign and malignant breast tumors using predictive modelling and analysis in machine learning.

Data Preprocessing

Image Acquisition: Collect the digital breast cancer images from the dataset.

Grayscale Conversion (if applicable): Convert the images to grayscale to reduce complexity, unless they need to be processed in color.

Noise Reduction: Apply filters to reduce noise and improve image quality, which can affect normalization effectiveness.

Min-Max Normalization:

- Identify the minimum and maximum pixel values across the dataset.
- Apply the Min-Max normalization formula to each pixel in the image:

$$X_{new} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

This scales all pixel values to a range between 0 and 1.

Data Partitioning: Divide the normalized images into training, validation, and test sets.

Augmentation (optional): To enhance the dataset and prevent overfitting, apply techniques like rotation, zoom, or horizontal flipping.

This approach ensures that the input to the classification model is standardized, which is crucial for achieving reliable and accurate diagnostic predictions.

CNN Architecture

To implement a Convolutional Neural Network (CNN) for breast cancer image classification, follow these general steps using a typical deep learning framework like TensorFlow with Keras:

Data Preparation: Load and preprocess images (e.g., normalization, resizing) and split them into training, validation, and testing datasets.

Model Building:

- Input Layer: Define the input layer to accept the shape of preprocessed images.
- Convolutional Layers: Add convolutional layers

with activation functions (usually ReLU) to extract features.

- Pooling Layers: Include pooling layers (usually max pooling) to reduce dimensionality.
- Flatten Layer: Flatten the output from convolutional networks to feed into fully connected layers.
- Fully Connected Layers: Add one or more dense layers for classification.
- Output Layer: Use a softmax activation layer for multi-class classification (benign vs malignant).

Compilation: Compile the model with an appropriate optimizer (like Adam), loss function (like categorical crossentropy), and metrics (like accuracy).

Training: Train the model using the training data with specified epochs and batch size, and use validation data to tune the hyperparameters.

Evaluation: Assess the model performance using the test dataset to check its generalization.

Prediction: Deploy the model to make predictions on new breast cancer images.

4. Experiments and Results

4.1. Experimental Setup

Hardware

For breast cancer image classification using a Convolutional Neural Network (CNN) with Min-Max Normalization, you'll need adequate hardware to handle the computational requirements effectively. Here are some recommended hardware specifications:

CPU: A multi-core processor (Intel i5, i7, or Xeon, AMD Ryzen or Threadripper) for efficient data handling and preprocessing.

GPU: A dedicated NVIDIA GPU (such as the GTX 1080, RTX 2060, or better) with CUDA support to accelerate the training of deep learning models. The more memory, the better (8GB VRAM minimum).

RAM: At least 16GB of RAM, though 32GB or more is recommended for handling large datasets and intensive computing tasks.

Storage: SSD (Solid State Drive) with at least 512GB of space for faster data access and storage of large image datasets and models.

Cooling System: Adequate cooling solutions to prevent thermal throttling during extended training sessions.

This setup should provide a solid foundation for training and deploying CNNs efficiently for medical image analysis tasks like breast cancer classification.

Software

For breast cancer image classification using a Convolutional Neural Network (CNN) with Min-Max Normalization, the following software components are essential:

Operating System: Windows, Linux, or macOS, depending on your preference and hardware compatibility.

Programming Language: Python is widely used due to its extensive support for data science and machine learning libraries.

Deep Learning Frameworks: TensorFlow or PyTorch are popular choices that provide robust tools and libraries for designing, training, and deploying CNNs.

Libraries and APIs:

- NumPy and Pandas for data manipulation.
- OpenCV for image processing.
- Matplotlib or another visualization tool for displaying images and graphs.
- Scikit-learn for additional machine learning functionalities.
- Keras (integrated within TensorFlow) for building and training deep learning models more conveniently.

Development Environment: Jupyter Notebook or PyCharm for coding, testing, and debugging your model in an interactive environment.

This software stack will enable the development and execution of effective image classification models for medical imaging tasks.

Metrics

For evaluating the performance of a Convolutional Neural Network (CNN) for breast cancer image classification using Min-Max Normalization, the following metrics are commonly used:

Accuracy: Measures the overall correctness of the model, calculated as the ratio of correctly predicted observations to the total observations.

Precision: Indicates the proportion of positive identifications that were actually correct, particularly important when the cost of a false positive is high.

Recall (Sensitivity): Measures the proportion of actual positives that were correctly identified, crucial for medical diagnostics to ensure all potential cases are considered.

F1 Score: The harmonic mean of precision and recall, providing a balance between them and is especially useful when the class distribution is uneven.

AUC-ROC Curve: The area under the receiver operating characteristic curve; this metric is useful for evaluating the performance at various threshold settings.

These metrics provide a comprehensive evaluation of the model's performance, considering aspects like model accuracy, the balance between sensitivity and specificity, and the ability to handle imbalanced datasets effectively.

4.2. Results and Analysis

The feature matrix, which is supplied as input to the model, is created using the retrieved features. The pre-trained model is utilized initially, and transfer learning is employed to adjust the hyperparameters. The input layer of the Visual Geometry Group16 (VGG16) model is set up for three channels. Since it's grayscale in our situation, the value from one plane is replicated to the other two.

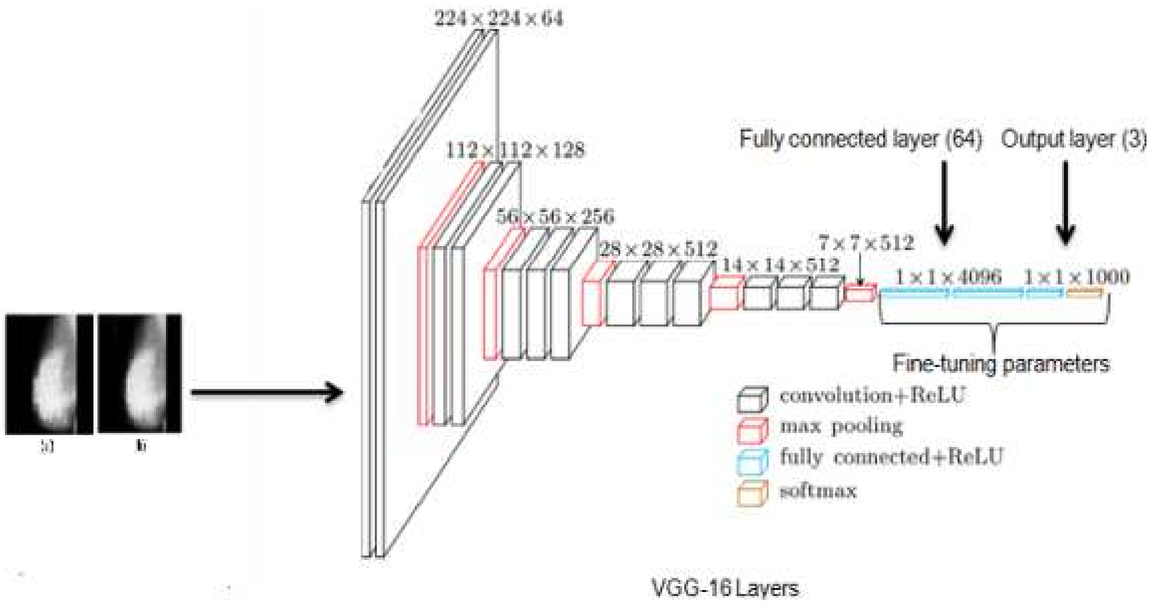


Figure 1: VGG 16 architecture details

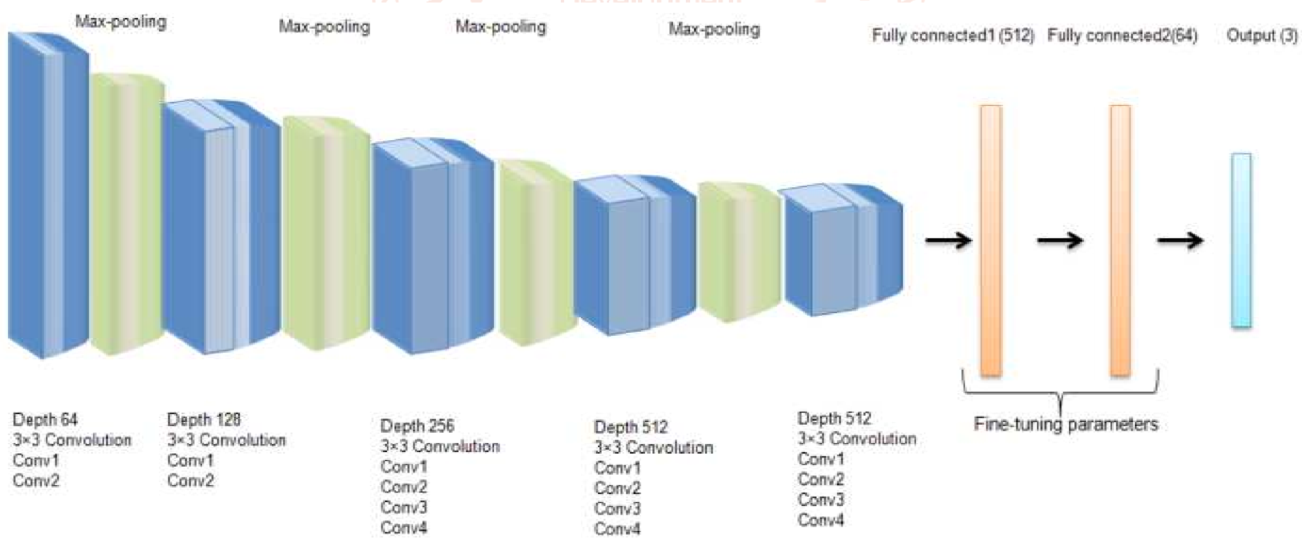


Figure 2: VGG 19 architecture details

Figures 1, 2 and 3 depict the architecture for VGG16, VGG19, and the constructed CNN architecture. A Figure 4, 5 & 6, illustrates the accuracy and loss with respect to epoch for the models VGG16, VGG19 and constructed CNN architecture. It is imperative from tables (2, 3 &4) the highest accuracy of 93 % using the constructed CNN architecture.

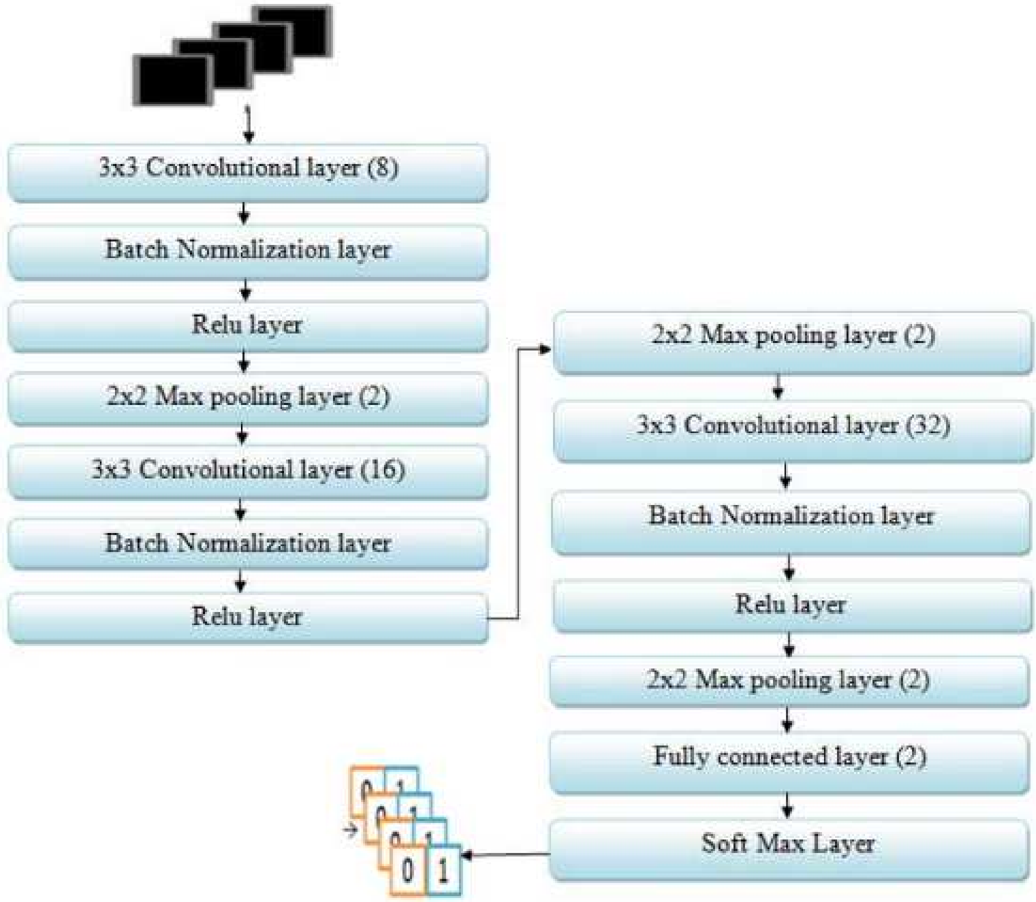
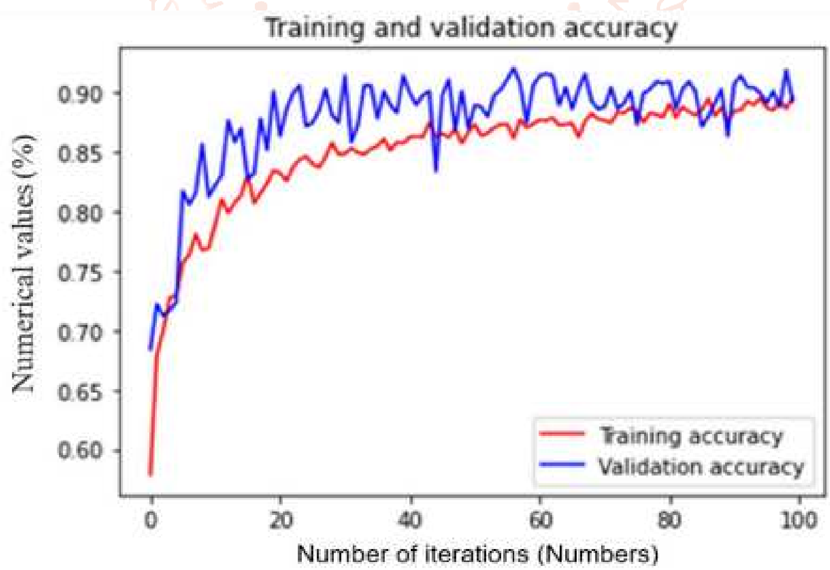
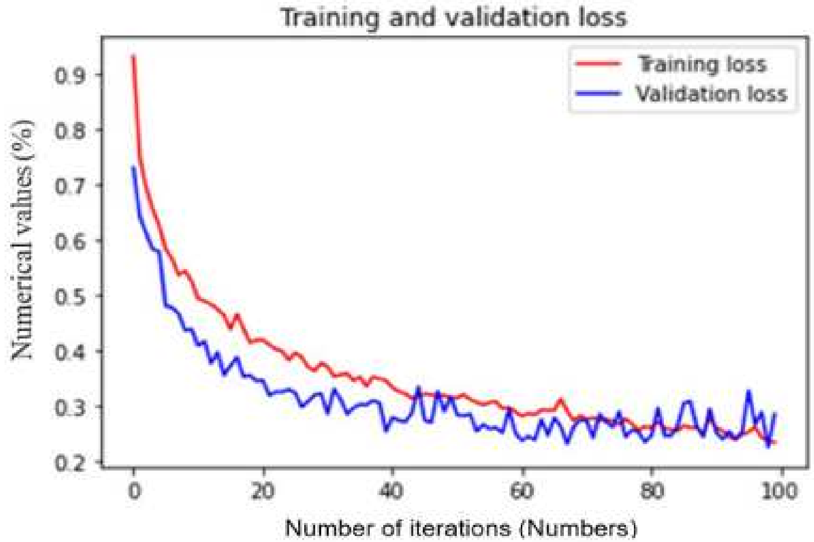


Figure 3: Constructed CNN architecture

The constructed CNN architecture comprises fourteen layers. Convolution, Batch normalization, Relu activation, max-pooling operation repeated three times followed by fully connected layer and a SoftMax layer.

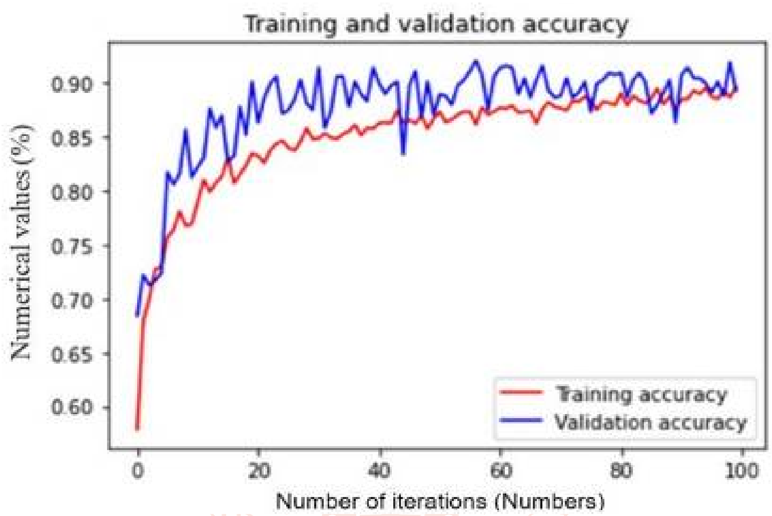


(a)

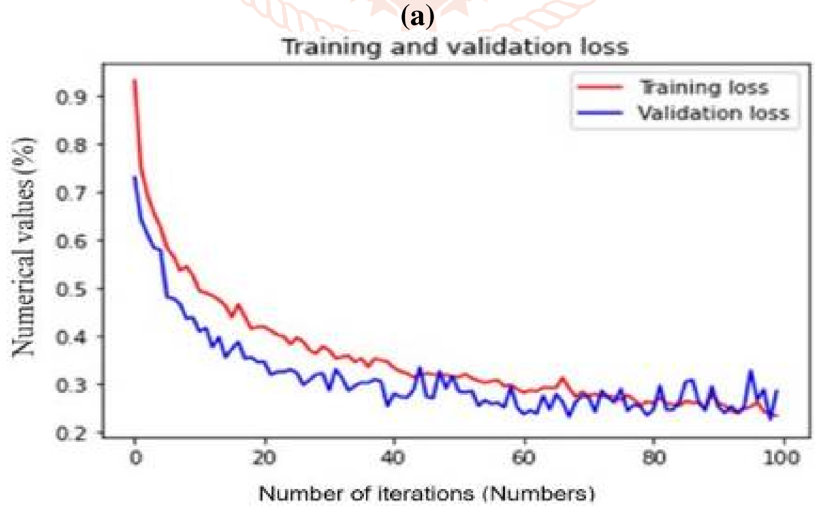


(b)

Figure 4: (a) & (b) Accuracy and loss graph obtained for VGG16 architecture (X axis unit in numbers, Y axis unit in percentage)

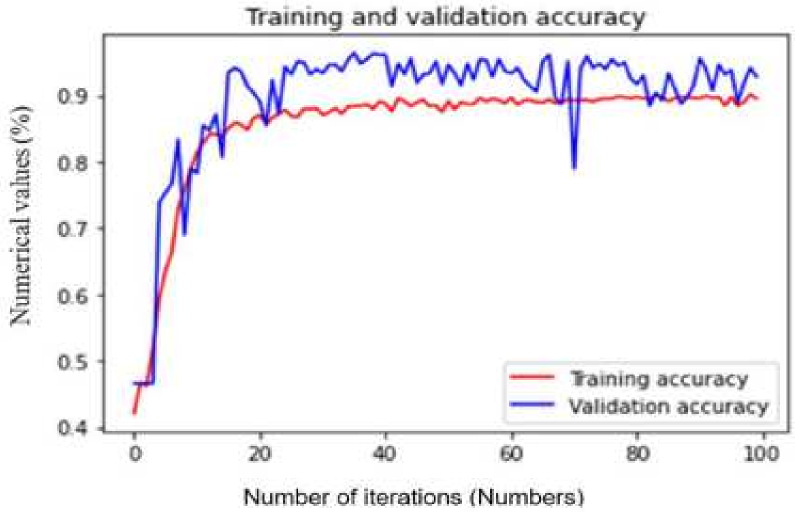


(a)

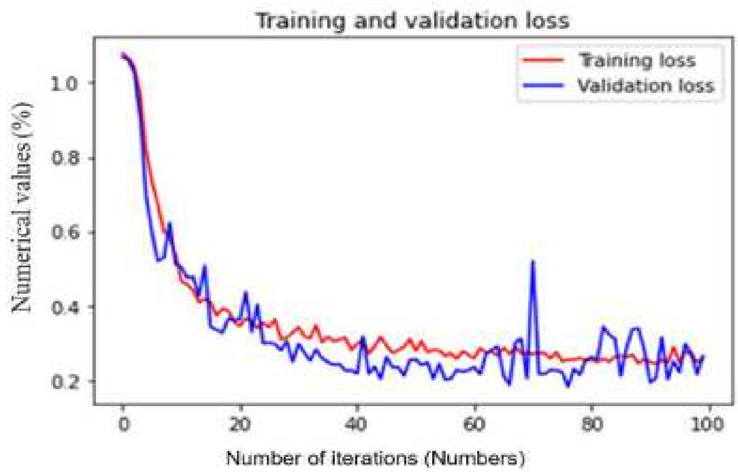


(b)

Figure 5: (a) and (b) Accuracy and loss graph obtained by using VGG19 architecture by transfer learning (X axis unit in numbers, Y axis unit in percentage)



(a)



(b)

Figure 6: Accuracy and loss graph obtained by using the constructed CNN architecture by transfer learning (X axis unit in numbers, Y axis unit in percentage)

Table 1: Results obtained with VGG16, VGG19, CNN-MMO (Proposed) architecture

Parameters	VGG16	VGG19	CNN-MMO (Proposed)
Precision	0.904	0.907	0.94
Recall	0.884	0.89	0.92
F1-Score	0.896	0.898	0.93
Accuracy	0.899	0.897	0.93

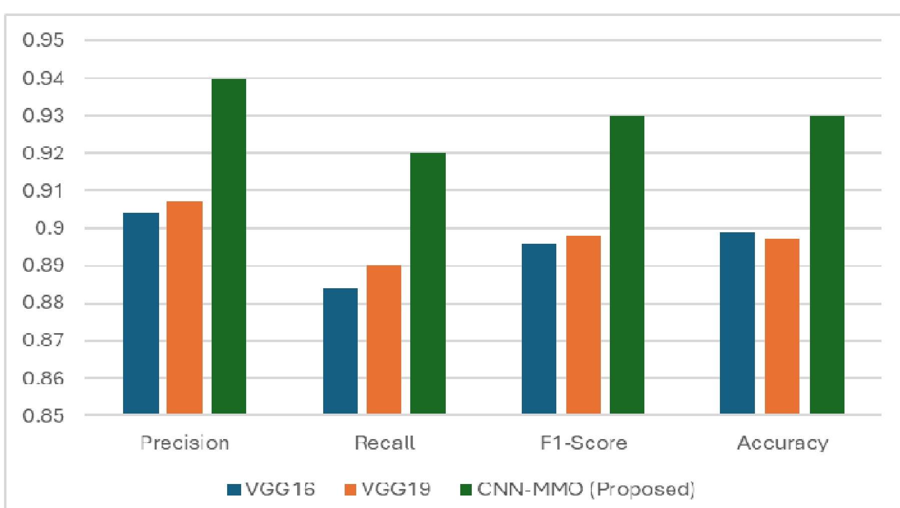


Figure 7: Graphical Analysis obtained with VGG16, VGG19, CNN-MMO (Proposed) architecture

Table 1, illustrates the comparison results obtained with VGG16, VGG19 and CNN-MMO architectures using transfer learning method. Figure 7 depicts the graphical illustrations of comparison results obtained with VGG16, VGG19 and CNN-MMO architectures using transfer learning.

5. Conclusion and Future Work

This study suggests a fresh approach to breast cancer early detection. Min-max optimization is used for background removal, pectoral muscle removal, contrast enhancement, and noise reduction in the first step of pre-processing. The application of wavelet and curvelet transformations comes next. The transform coefficients are used to extract features. The pretrained model and built-in convolutional neural network use this feature matrix as input. The idea of transfer learning investigates pre-trained models like VGG16 and VGG19. There are fourteen strata in the architecture. In order to achieve high classification accuracy, hyper-parameters are then adjusted. A maximum accuracy of 93% is attained using the convolutional neural network architecture that was created. The acquired accuracy surpasses the literature's state-of-the-art approach.

Future research focuses on the crucial need for a new technological development or improvements in medical image processing in order to treat cancer tissues a little bit more quickly. In order for clinicians to provide a better-informed diagnosis, these advancements must also be focused on producing more precise outcomes. However, by enhancing the associated highlighted features, this proposed system design might be improved in the future.

- Investigate different normalization and regularization strategies (e.g., instance normalization, group normalization) in combination with CNNs.
- Extend the architecture to a multi-task setting where tumor localization and type classification are performed simultaneously.
- Explore domain adaptation or transfer learning approaches to handle multi-institutional data, addressing dataset shift problems.
- Incorporate explainable AI (XAI) techniques for better interpretability and acceptance in clinical workflows.

References

[1] Pérez-Núñez, J., Rodríguez, C., Vásquez-Serpa, L., & Navarro, C. (2024). The Challenge of Deep Learning for the Prevention and Automatic Diagnosis of Breast Cancer: A Systematic review. Diagnostics, 14(24), 2896. https://doi.org/10.3390/diagnostics14242896

[2] Pranolo, A., Setyaputri, F. U., Paramarta, A. K. I., Triono, A. P. P., Fadhilla, A. F., Akbari, A. K. G., Utama, A. B. P., Wibawa, A. P., & Uriu, W. (2024). Enhanced multivariate time series analysis using LSTM: A comparative study of Min-Max and Z-Score normalization techniques. ILKOM Jurnal Ilmiah, 16(2), 210–220. https://doi.org/10.33096/ilkom.v16i2.2333.210-220

[3] Jabeen, K., Khan, M. A., Hameed, M. A., Alqahtani, O., Alouane, M. T., & Masood, A. (2024). A novel fusion framework of deep bottleneck residual convolutional neural network for breast cancer classification from mammogram images. Frontiers in Oncology, 14. https://doi.org/10.3389/fonc.2024.1347856

[4] Laska, R. R., & Yolanda, A. M. (2024). A comparative study of Z-Score and Min-Max normalization for rainfall classification in Pekanbaru. Journal of Data Sciences, 2024(1). https://doi.org/10.61453/jods.v2024no04

[5] Adam, R., Dell'Aquila, K., Hodges, L., Maldjian, T., & Duong, T. Q. (2023). Deep learning applications to breast cancer detection by magnetic resonance imaging: a literature review. Breast Cancer Research, 25(1). https://doi.org/10.1186/s13058-023-01687-4

[6] Alshayeji, M. H., & Al-Buloushi, J. (2023). Breast cancer classification using concatenated Triple Convolutional Neural Networks model. Big Data and Cognitive Computing, 7(3), 142. https://doi.org/10.3390/bdcc7030142

[7] Ali, M. D., Saleem, A., Elahi, H., Khan, M. A., Khan, M. I., Yaqoob, M. M., Khattak, U. F., & Al-Rasheed, A. (2023). Breast Cancer Classification through Meta-Learning Ensemble Technique Using Convolution Neural Networks. Diagnostics, 13(13), 2242. https://doi.org/10.3390/diagnostics13132242

[8] Shanmugavadivu, P., Kanimozhi, G., Dhamodharan, S., & Nithya, A. (2023). Breast cancer classification using optimized convolutional neural networks on enhanced segmented mammograms. AIP Conference Proceedings, 2929, 050004. https://doi.org/10.1063/5.0166173

[9] Yusoff, M., Haryanto, T., Suhartanto, H., Mustafa, W. A., Zain, J. M., & Kusmardi, K. (2023). Accuracy Analysis of deep learning methods in breast Cancer Classification: A

Structured review. Diagnostics, 13(4), 683. https://doi.org/10.3390/diagnostics13040683

[10] Jabeen, K., Khan, M. A., Alhaisoni, M., Tariq, U., Zhang, Y., Hamza, A., Mickus, A., & Damaševičius, R. (2022). Breast Cancer Classification from Ultrasound Images Using Probability-Based Optimal Deep Learning Feature Fusion. Sensors, 22(3), 807. https://doi.org/10.3390/s22030807

[11] Henderi, H. (2021). Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer. IJIIS International Journal of Informatics and Information Systems, 4(1), 13–20. https://doi.org/10.47738/ijiis.v4i1.73.